

Claims

- [c1] A method for improving compression of data, comprising:
arranging the data on a mixed format physical layout;
dividing the data on a mixed format physical layout into fixed-sized fields and variable sized fields; and
compressing the data of the variable sized fields and the fixed-sized fields.
- [c2] The method defined in claim 1, further comprising:
storing sizes of the fixed-sized fields in a data dictionary;
storing frequency of the data in the fixed-sized fields and the variable-sized fields in the data dictionary; and
storing information common to all records in the fixed-sized fields and variable sized fields in the data dictionary.
- [c3] The method defined in claim 1, wherein the fixed-sized fields further comprise field values.
- [c4] The method defined in claim 1, wherein the fixed-sized fields further comprise field offsets.
- [c5] The method defined in claim 1, wherein the fixed-sized fields further comprise pointers into the data dictionary.
- [c6] The method for compressing the data in the fixed-sized fields as defined in claim 3, further comprising:
storing a field value of the fixed-sized field in an additional variable-sized field;
coding the value of the fixed-sized field as of a field offset by pointing the field offset to the additional variable-sized field.
- [c7] The method for compressing the data in the variable-sized fields as defined in claim 3, further comprising:
storing frequently occurring long values of the fields in the data dictionary;
coding a value of the variable-sized field as of the field offset by pointing the field offset into the data dictionary, wherein the value of the variable-sized field is a redundant value.

[c8] The method for compressing the data in the variable-sized fields as defined in claim 5, further comprising:
coding a value of the variable-sized field as of the field offset by encoding the field offset into a record, wherein the value of the variable-sized field is a non-redundant value.

[c9] The method for compressing the data in the variable-sized fields as defined in claim 3, further comprising:
storing frequently occurring long values of the fields in a second data dictionary, wherein the second data dictionary is larger than the data dictionary;
coding a value of the variable-sized field as of the field value by pointing the field value into the second data dictionary, wherein the field offset is not large enough for the second data dictionary.

[c10] A method for improving compression of data, comprising: arranging the data on a mixed format layout, wherein the data comprises of a group of correlated fields;
dividing the data on a mixed format physical layout into fixed-sized fields and variable-sized fields; and
compressing the data of the variable-sized fields and the fixed-sized fields.

[c11] The method defined in claim 10, further comprising:
storing sizes of the fixed-sized fields in a data dictionary; storing frequency of the data in the fixed-sized fields and the variable-sized fields in the data dictionary;
storing information common to all records in the fixed-sized fields and variable sized fields in the data dictionary.

[c12] The method defined in claim 10, wherein the fixed-sized fields comprise of field values.

[c13] The method defined in claim 10, wherein the fixed-sized fields comprise of field offsets.

[c14] The method defined in claim 10, wherein the fixed-sized fields comprise of pointers into the data dictionary.

[c15] The method for compressing the data as defined in claim 12, further comprising:
storing frequently occurring values for the group of correlated fields in the data dictionary; and
coding a frequently occurring value for the group by pointing the field offset, belonging to the group, to the data dictionary.

[c16] The method for compressing the data as defined in claim 15, further comprising:
coding an infrequently occurring value for the group, wherein the field offset, belonging to the group, points to the record in the field.

[c17] The method for retrieving compressed data, comprising:
receiving a request for data in a compressed database;
receiving compressed database on a mixed format physical layout responsive to the request, wherein the mixed format physical layout comprises of fixed fields and variable fields;
searching for a value in the fixed fields corresponding to the request for data;
retrieving the value in the fixed fields corresponding to the requested data.

[c18] The method defined in claim 17, wherein the retrieving step further comprises:
retrieving a dictionary entry if the value of the fixed field comprises of a dictionary pointer;
retrieving a value starting from a field offset if the value of the fixed field comprises of the field offset; and
retrieving a same field from that record, if the value of the fixed field comprises of a record offset.

[c19] An apparatus for improving compression of data, comprising:
means for arranging the data on a mixed format physical layout;
means for dividing the data on a mixed format physical layout into fixed-sized fields and variable sized fields; and
means for compressing the data of the variable sized fields and the fixed-sized fields.

[c20] An apparatus for retrieving a compressed data, comprising:
means for receiving a request for decompressing a requested compressed data;
means for receiving the compressed data on a mixed format physical layout
responsive to the request, wherein the mixed format physical layout comprises
of fixed fields and variable fields;
searching for a value in the fixed fields;
means for retrieving the value in the fixed fields corresponding to the requested
compressed data.

[c21] A compressible computer medium, comprising a plurality of instructions to
cause a computer to perform the steps of:
arranging the data on a mixed format physical layout;
dividing the data on a mixed format physical layout into fixed-sized fields and
variable sized fields; and
compressing the data of the variable sized fields and the fixed-sized fields.

[c22] The compressible computer medium according to claim 21, wherein the
instructions further cause the computer to perform the steps of:
storing sizes of the fixed-sized fields in a data dictionary;
storing frequency of the data in the fixed-sized fields and the variable-sized
fields in the data dictionary;
storing information common to all records in the fixed-sized fields and variable
sized fields in the data dictionary.

[c23] The compressible computer medium of claim 21, wherein the fixed-sized fields
comprise of field values.

[c24] The compressible computer medium of claim 21, wherein the fixed-sized fields
comprise of field offsets.

[c25] The compressible computer medium of claim 22, wherein the fixed-sized fields
comprise of pointers into the data dictionary.

[c26] The compressible computer medium according to claim 23, wherein the
instructions further cause the computer to perform the steps of:
storing a value of the fixed-sized field in an additional variable-sized field;

coding the value of the fix-sized field as of a field offset by pointing the field offset to the additional variable-sized field.

[c27] The compressible computer medium according to claim 23, wherein the instructions further cause the computer to perform the steps of:
storing frequently occurring long values of the fields in the data dictionary;
coding a value of the variable-sized field as of the field offset by pointing the field offset into the data dictionary, wherein the value of the variable-sized field is a redundant value.

[c28] The compressible computer medium according to claim 25, wherein the instructions further cause the computer to perform the steps of:
coding a value of the variable-sized field as of the field offset by encoding the field offset into a record, wherein the value of the variable-sized field is a non-redundant value.

[c29] The compressible computer medium according to claim 23, wherein the instructions further cause the computer to perform the steps of:
storing frequently occurring long values of the fields in a second data dictionary, wherein the second data dictionary is larger than the data dictionary;
coding a value of the variable-sized field as of the field value by pointing the field value into the second data dictionary, wherein the field offset is not large enough for the second data dictionary.